ORIGINAL

EX PARTE OR LATE FILED

AT&T

Frank S. Simone
Government Affairs Director

Suite 1000
1120 20th Street, N.W.
Washington, DC 20036
202 457-2321
FAX 202 457-2545
EMAIL fsimone@att.com

July 20, 1999

RECEIVED

JUL 2 0 1999

FEDERAL COMMUNICATIONS COMMISSION
OFFICE OF THE SECRETARY

Ms. Magalie Roman Salas, Secretary
Federal Communications Commission
445 Twelfth Street, S. W. – Room TWB-204
Washington, D. C. 20554

Re: *Ex parte*, CC Docket No. 98-56, Performance Measurements and Reporting Requirements for Operations Support Systems, Interconnection, and Operator Services and Directory Assistance

Dear Ms. Roman Salas:

On Tuesday, July 20, 1999, the attached document, an analysis of BellSouth OSS performance using BellSouth provided data, was delivered to Daniel Shiman of Common Carrier Bureau's Policy and Program Planning Division. Please include a copy of this Notice in the record of the above-captioned proceeding.

Two copies of this Notice are being submitted to the Secretary of the FCC in accordance with Section 1.1206 (b) of the Commission's rules.

Sincerely,

F S Simone

ATTACHMENT

cc: M. Pryor
D. Shiman
C. Pabo
J. Stanley
E. Einhorn

No. of Copies rec'd 0+2
List A B C D E

Notes On Some Analyses Of BellSouth Data

Colin Mallows, AT&T Labs - Research
June 16, 1999

<u>Contents</u>

Notes On Some Analyses Of BellSouth Data

Colin Mallows, AT&T Labs - Research
June 16, 1999

1. Introduction and summary.

BellSouth gave me (on March 5, 1999) a JAZ disc containing filtered Order Completion Interval data for both BST and the aggregated CLECs for August through December 1998. On March 22 I received another JAZ disc, containing Maintenance Average Duration data also. A third JAZ disc containing Missed Repairs and Missed Installations data arrived later.

My purpose in this study is to demonstrate that:

(a) the approach advocated by LCUG is feasible,
(b) it gives different results from those of BellSouth,
(c) there are systematic effects in the data that are not treated appropriately by the BellSouth approach.

I have had the benefit of some comments from Ernst and Young on earlier versions of this report.

In performing these analyses, I have found it necessary to develop novel methods for aggregating the results from many individual cells into an overall statistic. Appendix A discusses some of the issues that arise.

The present results apply only to the August and September OCI data, the August MAD data, and the Missed Repairs data for August through December 1998. However they suggest the following general conclusions.

1. The LCUG approach is feasible, and gives detailed results at the wire-center/class (= "cell") level. If permutation calculations are employed, it can be used even for the smallest small sample sizes (down to one observation from each of BST and CLEC).

2. If the results from individual cells are aggregated in the way I suggest, the results will be different from those obtained by E&Y, because (i) their method allows good results in one cell to cancel out bad performance in another, and (ii) their method allows the variability between cells to enter into the scale of the reference distribution to which the observed differences are compared.

3. For these data, and possibly in general, if the sample sizes are so small as to make a permutation calculation necessary, the difference between the permutation method

using LCUGZ and the permutation method using the classical "Pooled" Z is small. This suggests that we can use the much faster PooledZ in permutation calculations.

4. Permutation calculations are unnecessary for "counted" data such as Missed Repair Appointments, because there is a standard formula for this case using the Hypergeometric distribution.

5. For measured variables, permutation calculations can be avoided by either
   (a) aggregating the data to make the sample sizes large enough, or
   (b) transforming the raw data, for example by taking square roots.

6. In the Missed Repairs data there are systematic differences among the wire centers, which are treated as random effects in the BellSouth approach. The effect is that in their approach all these differences contribute to the estimate of variance, thus deflating the value of the overall statistic.

2. Order Completion Interval.

2.1 The August OCI data

In this section I demonstrate that the LCUG approach is feasible for data of this kind, and gives results different from those obtained by Ernst and Young for BellSouth. My work has built on that of Ernst and Young. In particular, I have disaggregated to the wire-center (w/c) level, (221 w/cs are represented in the August OCI data), and have not attempted to improve on their classification of the observations. Ernst and Young classified the OCI data according to these five variables:

Order type:	3 levels:	Change, New, Transfer (there were no "Prewire" cases in these data)
Dispatched:	2 levels:	Dispatched, Non-dispatched
Residence:	3 levels:	Residence, Business, Special
Circuits:	2 levels:	<=10, >10
Day of month:	2 levels:	<=15, >15

This makes 72 classes, for a total of 221*72=15912 possible w/c-class combinations. I call these "cells". Every occupied cell has both BST and CLEC data. The number of observations in the cells varies very widely; however each cell contains a statistically valid sample for analysis. The following table shows how the data are distributed over the cells:

Range of nCLEC	*Proportion of cells*
100 -	1.5 %
30 – 99	7.8 %
10 – 29	15.8 %
1 - 9	74.9 %

2.2 The LCUG approach

The published LCUG documents say nothing about disaggregation. In those documents, it was assumed that the method would be applied in situations where like was being compared with like. However, Ernst and Young have since pointed out that it is important to disaggregate the data, because otherwise biases can be introduced that give the illusion of discrimination even when perfect parity exists in every cell. The LCUG approach can be applied within each disaggregated cell, but this will give a large number of Z-values, and it is not clear how these should be assembled into one overall statistic. There is as yet no "LCUG method" for dealing with reaggregation. Below, I suggest a way of combining the individual Z-values into a single overall index. Ernst and Young's "Adjusted Modified LCUGZ 2" statistic can be regarded as applying a different method of aggregation. Appendix A discusses this issue.

My main purpose in the study of the OCI data has been to study whether the LCUG approach is feasible, and to explore how far it is necessary to use permutation calculations. The basic LCUG proposal is to use the "modified Z" approach on comparable quantities. A first step is to calculate LCUGZ for the cells with nCLEC at least 100. We find that 38% of these values of LCUGZ are less than –2, and 29% are less than –3. This suggests that parity of service is not being provided, since under normal theory the expectation of the proportion of cells with $Z<-2$ is 2.3%, and for $Z<-3$ it is 0.14%.

However we cannot rely on normal theory, because the data are far from normal, being in fact discrete, taking only integer values, and very skew (long-tailed to the right). We can perform permutation calculations to obtain more accurate results. Appendix B explains how this calculation gives an adjusted value of Z. Figure 1 plots LCUGZ against values obtained by the permutation method, using 1000 permutations. Evidently the method is working well, with the two values agreeing very closely, except for the very large and very small values. There are several cells for which the permutation Z is not as extreme as the LCUGZ; this happens because in the most extreme possible case, where the observed value of LCUGZ is more extreme than each of the 1000 randomly-generated permutation values, we get a value of –3.09 for permLCUGZ. If we used more permutations, we could get more extreme values, and Figure 1 would look straight further out on the tails.

Since the computation of LCUGZ is time-consuming, it would be helpful if we could replace it by the standard Pooled Z in the permutation calculation. (Using PooledZ is much faster because for each permutation we need only compute the sum of the selected CLEC observations.) Figure 2 compares the values of permLCUGZ with those of permPooledZ. We see that the two statistics agree very well. The following Table shows the numbers of cells for which the computed values of Z are less than –2 and –3, for each of the three methods of computation, namely the simple LCUG formula, the permutation-adjusted version of LCUGZ and the permutation version of PooledZ.

Results of calculations by three methods

nCLEC	%cells	method	Z<-2 observed	Z<-2 expected	Z<-3 observed	Z<-3 expected
100 -	1.5%	LCUGZ	38%	2.3%	29%	0.13%
		permLCUGZ	33%		24%	
		permPooledZ	38%		24%	
30 – 99	7.8%	LCUGZ	26%	2.3%	17%	0.13%
		permLCUGZ	23%		14%	
		permPooledZ	26%		15%	
10 – 29	15.8%	LCUGZ	17%	2.3%	10%	0.13%
		permLCUGZ	13%		4%	
		permPooledZ	14%		5%	
1 – 9 (and nBST >9)	53.9%	permPooledZ	7.6%	2.3%	31%	0.13%
1 – 9 (pooled over w/c)	0.9%	LCUGZ	31%	2.3%	31%	0.13%
		permLCUGZ	23%		23%	
		permPooledZ	23%		23%	

In the cells with nCLEC between 30 and 99, we again find evidence of violations of parity. Figures 3 and 4 are similar to Figures 1 and 2. Again we see that we can replace the permutation-adjusted version of LCUGZ by the permutation-adjusted version of PooledZ. However Figure 3 shows that there are systematic differences between LCUGZ and the permutation versions. Evidently a sample size of 30 is barely enough for normal theory to be acceptable, for these data.

Figures 5 and 6 give similar results for the cells with nCLEC between 10 and 29. Now the normal-theory LCUGZ calculation is clearly not adequate, but the two permutation methods agree well.

These calculations suggest that for the OCI data we can use LCUGZ whenever nCLEC is at least 100, and need to use a more accurate method when nCLEC is smaller than 100. The permutation method using PooledZ seems adequate.

74.9% of the occupied cells have fewer than 10 CLEC observations. These represent 33 different cell-types, and all 221 wire-centers. In 72% of these cells there are at least 10 BST observations. I have applied the permutation method to these, using PooledZ. The results appear in the Table above.

We are left with 21% of the cells that each have fewer than 10 CLEC and fewer than 10 BST observations. For 93% of these cells, representing 13 different types, I have pooled over wire-centers. (This aggregation may not be meaningful.) The results appear in the Table above. The remaining 1.6% of the occupied cells, containing less than 1/4000 of the observations, represent 14 different types. It would be possible to use the permutation method even for these very small cells, but I have not done this for the OCI data.

For these data, it is evidently necessary to use the permutation method whenever the number of CLEC observations in a cell falls below 100. We could avoid the necessity for permutation calculations if we could agree on meaningful aggregations over w/cs. Once set up, such an aggregation scheme could be used for several months, and revised only when the mix of activity changed significantly. It seems plausible that such a scheme could be used for each of several different measures, since activity in one measure is likely to be correlated with activity in another.

2.3 Transforming the data.

An alternative way of avoiding permutation calculations is to transform the data, for example by taking square roots. This could be implemented by routinely storing OCI and sqrt(OCI) (instead of OCI and OCI^2), so that with only minor variations the same processing steps as are used now would provide the mean and variance of sqrt(OCI) within each cell. Sqrt(OCI) has a distribution that is much closer to normal than OCI itself, and the simple normal-theory LCUG method should work well for cells with nCLEC smaller than 100. I repeated all the above analysis using sqrt(OCI) values, finding that the simple LCUG calculation agrees with the results from the permutation method more closely than when the raw data is used.

2.4 An overall statistic.

We have defined a large number of cells (including the "pooled" ones) and need to combine their Z-values to arrive at an overall indicator of compliance. Appendix A discusses how this might be done. There are many possibilities. For example, we can simply count, as above, how many Z's fall below –2, and compare this with the number to be expected, assuming parity. Or we could use –3, or any other value. We should not simply average the Z-values, because that would allow positive values to cancel out negative ones. Also it seems clear that some weighting should be used, because a Z-value that is based on hundreds of observations is worth more than one based on only 20. I suggest the following procedure. First, replace all positive Z's by zero, thus getting

$$Z^* = \text{truncZ} = \min(0,Z)$$

Form a weighted average of the resulting quantities, using as weights

$$w = 1/\text{sqrt}\,(1/\text{nCLEC} + 1/\text{nBST})$$

Under the parity hypothesis, this weighted average will have an expected value and variance of

$$M = -\text{sum}(w)/\text{sqrt}(2\text{pi}), \qquad V = \text{sum}(w\text{^}2)(1/2 - 1/2\text{pi})$$

respectively, so we can define the statistic

$$\text{Zagg} = (\text{sum}(wZ^*) - M)/\text{sqrt}\,(V)$$

which should have a standard normal distribution under the parity hypothesis. Appendix A explains how M and V can be obtained. For the August OCI data, we get the value

$$\text{overallZ} = -11.70$$

which is astronomically significant. (The tail probability is about 10^-32). For comparison, Ernst and Young apply six different methods; (or twelve if we count the "adjusting to BST" methods). They obtain the results

Method	test statistic	Method	test statistic
"LCUG"	-9.85	REPL	-4.46
"POOL"	-10.33	REPL ADJ	-4.11
		JACK	-4.03
Present method	-11.70	JACK ADJ	-3.72

(Their methods "LCUG" and "POOL" are different from the present method.) We see that our method makes the overall statistic more than twice as large as the Ernst and Young methods.

2.5 The September OCI data

Similar calculations have been performed on the September data, except that I have not dealt with the cells that have both nCLEC and nBST smaller than 10. The results appear below.

September OCI results:

nCLEC	%cells	method	Z<-2 observed	Z<-2 expected	Z<-3 observed	Z<-3 expected
100 -	2.1%	LCUGZ	41%	2.3%	36%	0.13%
		permLCUGZ	45%		36%	
		permpoolZ	50%		32%	
30 – 99	11.5%	LCUGZ	52%	2.3%	35%	0.13%
		permLCUGZ	53%		36%	
		permpoolZ	50%		28%	
10 – 29	21.1%	LCUGZ	38%	2.3%	25%	0.13%
		permLCUGZ	36%		21%	
		permpoolZ	32%		12%	
1 – 9	65.3%	LCUGZ	15.1%	2.3%	7.5%	0.13%
(and nBST >9)		permLCUGZ	10.9%		5.0%	
		permpoolz	7.5%		1.3%	

For the cells that appear in the table above, the overall statistic is:

$$\text{Zagg} = -18.46$$

The Ernst and Young statistics for this month are again about half as big as ours:

Method	test statistic	Method	test statistic
"LCUG"	-24.87	REPL	-9.85
"POOL"	-24.24	REPL ADJ	-10.21
		JACK	-9.87
Present method	-18.46	JACK ADJ	-10.24

3. Maintenance Average Duration.

3.1 The August MAD data

My purpose here is again to demonstrate that the LCUG approach can handle data of this kind. Here there are many fewer observations than for OCI. For these data, I have not attempted to deal with the w/c classification. I have ignored the observations for which the wire-center is undefined.

We cannot rely on normal theory for these data; we need to use permutation calculations throughout.

75% of the cells in the Designed category have a single CLEC observation, and fewer than 10 BST observations. We need some way of dealing with these very small cells. I have applied the method outlined above. Thus in each cell, we perform a permutation calculation to give an adjusted Z-value. Then these are truncated to form Z* values, and combined as suggested above. I obtain the following results:

Designed cases	Zagg = -4.65, suggesting a large bias in favor of BST,
Non-Designed	Zagg = -2.09, suggesting a smaller bias in favor of BST.

Ernst and Young obtained values ranging from –1.87 to –1.97 for the Non-Designed cases. I cannot find their overall Designed results.

4. Missed Repair Appointments

4.1 The Missed Repairs data.

This measure is a "counted" variable. Again, I demonstrate that the LCUG approach is feasible for data of this kind, and also that there are strong systematic differences between the cells, that are not treated appropriately by the BellSouth approach. For the Designed data, there are only two cells, Dispatched and Non-Dispatched. While the number of observations may seem to be large, the proportion of Missed Repairs is very small (only 3% of Dispatched, and less than 0.1% of Nondispatched cases in these five months), so that we cannot rely on the simple algebraic formula for a two-by-two table. We must use the exact Hypergeometric distribution. Appendix B explains how this can be used to give an adjusted value of Z. For the Designed data, the results are:

Designed Missed Repairs: adjusted values of Z

	Aug.	Sept.	Oct.	Nov.	Dec.
Disp.	-1.659	-2.415	0.669	0.683	-0.638
Nondisp.	0.030	-1.947	0.000	0.034	0.000

Evidently something happened in September, but otherwise there is little evidence of discrimination.

For the Non-Designed data, there are many more observations than for the Designed. These observations are disaggregated by wire-center and according to two attributes: Dispatched/Non- dispatched, and Residence/Business. I have applied two different methods to these counts.

4.2 Method 1

First, I ignored the wire-center information. This means we have just four cells (defined by the Dispatched/NonDispatched and Residence/Business attributes) in each time-period. I applied the (Tukey version of the) "arcsin" transformation to the resulting proportions. This transforms each proportion into a number between 0 and pi, and has the property that the variance of the transformed quantity is very nearly independent of the true proportion, and is very nearly 1/n where n is the number of observations. Taking the difference between the BST and CLEC transformed values within each cell, we get a 4 by 10 table of differences.

I applied a standard method, the Analysis of Variance, to decompose the variation among these differences into components representing

(1) the variation among the four cells
(2) the variation among the ten time-periods
(3) the interaction between the cell-classification and the time-periods.

The resulting Analysis of Variance table is:

Source	Sum of Squares	d.f.	Mean Square	P-value
Between Cells	70.791	3	23.597	3.10^-15
Between Time-periods	15.676	9	1.742	0.074
Interaction	36.643	27	1.357	0.102
Total	123.110	39		

If there were no systematic between-cell effects, the Between-Cell Mean Square should be approximately unity. It is clearly very much larger than this. A value this large will occur by chance only 3 times in 10^15, if there are indeed no differences among the cells. We deduce that there are large differences between the four cells. On the other hand, both the Between time-periods Mean Square and the Interaction Mean Square are approximately unity, (values as large as are observed here will occur by chance 7.4% and 10.2% of the time, if there are no differences among the time-periods and the only effects are the between-cell ones). This shows that the between-cells differences are consistent over time; there are no significant trends over time.

4.3 Method 2.

My second analysis takes the wire-center classification into account. I applied a variation of the BellSouth method to obtain, for each of the four cells and ten time-periods, the difference between the BST and CLEC proportions weighted by CLEC count. Appendix C explains how this was done, and derives the variance of the resulting weighted difference. Hence we get, in each of the 40 cells, a weighted Z statistic, which under the parity hypothesis should have mean zero and unit variance.

Applying the same Analysis of Variance method as above, we get the following results:

Source	Sum of Squares	d.f.	Mean Square	P-value
Between Cells	59.458	3	19.819	8.10^-13
Between Time-periods	13.498	9	1.500	0.141
Interaction	34.325	27	1.271	0.157
Total	107.281	39		

Now the Between-time-periods and Interaction mean squares are closer to unity, but the Between cells Mean square is again much larger than unity. We see that when wire-centers are taken into account in this way, there are still strong between-cells differences, but there are no significant differences over time, and the between-cell effects are consistent over time.

The fact that the between-cell sum of squares is reduced when the wire-center identification is taken into account shows that it is important to make this adjustment. A more direct assessment of the effect of the disaggregation by wire-centers can be obtained from the variance of the individual wire-center Z's. The empirical estimate of the variance of these Z's is 1.174 (with 1731 degrees of freedom), somewhat larger than unity. This shows that there are systematic differences among the wire-centers, which will have the effect of inflating the variance estimate used in the BellSouth approach by a factor of about sqrt(1.174) = 1.083.

5. Missed Installation Appointments

The MI data are classified just like the OCI data. Only one cell, namely New, Residential, NonDispatched, Less than 10 circuits, has enough data for a meaningful analysis to proceed. For this cell, there are just 31 wire-centers that have data for all ten half-months. I computed the standard Z statistic (as in Appendix B) for each of these 310 cases, and used the "symmetric" weights described in Appendix C. An Analysis of Variance calculation like that in Section 4.3 gave the following results. (Because the weights are not proportional, we get slightly different results by fitting the effects in two different orders).

Source	SSq	df	Meansq	P-value	Source	SSq	df	Meansq	P-value
W/c	89.504	30	2.984	8.10^(-8)	Time	22.358	9	2.484	.0078
Time|wc	20.435	9	2.271	0.015	W/c|Time	87.582	30	2.919	2.10^(-7)
Int'n	430.952	270	1.596	2.10^(-9)	Int'n	430.952	270	1.596	2.10^(-9)
Total	540.892	309			Total	540.892	309		

We see that there is very strong evidence of systematic differences among the wire-centers.

Appendix A. Disaggregation and reaggregation.

In testing for discriminatory behavior, as Ernst and Young have emphasized, it is important to disaggregate the observations on each SQM into cells that are small enough such that within each cell we are comparing like with like. Applying an analysis method to naively pooled data can introduce biases; data that exhibits perfect parity within each cell can appear to show discrimination when pooled.

Given disaggregated data, we can compute some measure of compliance within each cell. Ernst and Young simply compute the difference between the average BST and average CLEC measurements within each cell; they pay no attention to the dispersion within cells. The LCUG approach is to compute the LCUGZ statistic within each cell, possibly using permutation calculations. However this is done, the question arises, how should we reaggregate these measures into a single overall statistic? We want a method that has these properties:

(a) The method should provide a single overall index, which is on a standard scale.
(b) If the entries in the cells are exactly proportional, the reaggregated index should be very nearly the same as if we had not disaggregated.
(c) The contribution of each cell should depend on the number of observations in the cell.
(d) As far as possible, cancellation should not be allowed to occur.
(e) The index should be a continuous function of the observations.

The motivation for the requirement (d) is that, for example, BST should not be able to discriminate against CLEC Business customers, while avoiding detection by discriminating in favor of CLEC Residence customers. The motivation for (e) is that we do not want the final result to depend critically on minor details in the data. A small change in the data should induce a small change in the final result.

In the BellSouth "Replicate Variance" method, a weighted average of the within-cell differences is compared with a measure of scale that is obtained from the dispersion of these same differences. Their "Jackknife" method is a variant of the same approach. For the LCUG method, the individual Z-values that we get for each occupied cell are already on a standard scale. One approach (borrowed from earlier thinking on aggregation over measures) is simply to count how many of the within-cell Zs lie below 2 or 3 or some other number. This is what we did in the first analysis, of OCI data. This method satisfies requirements (a) and (d), but not (b), (c), or (e). Another possibility is to form a weighted average of the Z-scores. The weights should depend on the sample sizes. Requirement (c) will be satisfied if we take

aveZ = sum (wZ)/sqrt(sum w^2)

where

w = weight = 1/sqrt(1/nILEC + 1/nCLEC).

This aveZ statistic is on the standard normal scale, and satisfies all the requirements except (d). It is similar to the BellSouth "Modified Adjusted LCUG 2" statistic, which is of the same form, but which uses weights

$$w2 = w * sqrt(1 + nCLEC/nILEC) * sILEC$$

where sILEC^2 is the within-cell ILEC variance.

The "truncatedZ" proposal described in the main text satisfies all five desiderata. The final statistic

$$Zagg = (sum (W Z^*) - M)/sqrt(V)$$

is on the standard normal scale (provided there are a large number of cells). The truncation step reduces the possibility of cancellation (it does not remove it completely; but it does not allow BST to get extra credit by making any Z-score more positive). The measure is a continuous function of the observations. The values of M and V are

$$M = sum (W E(Z^*)) \qquad V = sum (W Var (Z^*))$$

and when the sample sizes are large,

$$E(Z^*) = 1/sqrt(2 pi) = 0.3989 \qquad Var (Z^*) = ½ - 1/(2 pi) = 0.3408$$

There is a minor technical difficulty in the calculation of M and V when the sample sizes are very small. In this case, if we obtain a value of Z by a permutation calculation, then Z will be only approximately normal. For example, if nBST = nCLEC = 2 and parity holds, and if the observations are in general position so that no ties occur, then the value of permZ is uniform on the six values.

Q(1/12)	Q(3/12)	Q(5/12)	Q(7/12)	Q(9/12)	Q(11/12)

Where Q is the inverse Normal function, i.e.

-1.3830	-0.6745	-0.2104	0.2104	0.6745	1.3830

Thus the truncated Z* is uniform on the six values

-1.3830	-0.6745	-0.2104	0	0	0

and the moments of Z* are

$$E(Z^*) = 0.3780 \qquad var(Z^*) = 0.2591$$

There is no difficulty in calculating these moments exactly as functions of nBST and nCLEC, even when there are ties in the data. Thus we can find the exact values for M and V in all cases.

The E&Y approach would allow some violations of parity to go undetected. For example, suppose it consistently occurs that in the first half of every month, the CLEC mean exceeds the BST mean, while in the second half the reverse is true. E&Y would regard this effect as part of the overall randomness in the system, while in my view there is systematic violation of parity for the first half of every month. If it is correct to disaggregate to the two-week level, we should look for departures from parity at that level.

Appendix B. Using a permutation calculation to compute an adjusted value of Z.

The permutation method calculates a large number of values of the chosen statistic, in our case LCUGZ, or PooledZ, by relabeling the observations as to BST or CLEC. In principle, we should compute all possible relabelings; in practice the number of possible relabelings is huge, and instead we use the computer to generate a pseudo-random sample of them. Suppose we decide to compute N such values. With the actual observed value Z0, this gives N+1 values of Z. Under the parity hypothesis, each of this set of values is equally likely. It follows that under parity, the rank R0 of Z0 in this set is equally likely to take any value in the range 1,...,N+1. To obtain a test with type-I error equal to p, we compute Rp=(N+1)*p, and claim violation of parity whenever R0 is one of the values 1,2,...,Rp. This is equivalent to claiming violation whenever Q((R0-1/2)/N)<=Q((Rp-1/2)/N), where Q is the inverse normal distribution function. (Q(p) is the value such that the probability is exactly p that a standard normal variable is less than (Q(p)). Under the parity hypothesis, the distribution of Q((R0-1/2)/N) is very close to a standard normal distribution; so Q((Rp-1/2/N) can be viewed as an appropriately adjusted critical value, and Q((R0-1/2)/N) itself can be regarded as an adjusted value of LCUGZ.

If the observations are integers, as in the OCI data, ties become a problem. These can be dealt with by taking R0 to be the average of its possible values, given the reference set of permutation Z-values. Thus we take R_0 = (# less) + (# equal)/2+1.

If the observations are counts, as for the Missed Repair Appointments data, in each cell we have a two-by-two table such as

	Missed	Non-missed	Total
CLEC	a	b	m
BST	c	d	n
Total	r	s	N

The standard formula

Z = (a - mr/N)/sqrt(mnrs/N^2(N-1)). (1)

gives a statistic that is approximately standard normal; the approximation is excellent if all the marginal totals are large. However when any of m,n,r,s is small, it is unnecessary to perform the permutations explicitly, because we have an exact formula for the probability of obtaining each possible value of a,b,c,d, given the margins m,n,r,s. (We assume that these marginal totals give no information regarding the presence or absence of parity). The formula is:

P(a,b,c,d | m,n,r,s) = m!n!r!s!/a!b!c!d!N!

Hence we can determine, for any observed value of a, the probability of getting a value as extreme as this, under the parity hypothesis, and this can be transformed into an equivalent value of Z. When the counts are very small, it helps to use a "continuity correction". Given an observed value a0, we define Z to be the value Z0 such that

Q(Z0) = P(a<a0|m,n,r,s)+1/2P(a=a0|m,n,r,s)

Appendix C. Using BST weights.

In the Missed Repairs data, we have (for each month) counts in each of eight cells, and many wire-centers. (Some of these cells are empty). Thus in each cell we have a two-by-two table like that shown in Appendix B. We can aggregate over wire-centers by forming the weighted average of the differences between the BST and CLEC proportions, using as weights the CLEC counts. This gives

$$A = \text{sum}(m(a/m - c/n))/\text{sum}(m)$$

We can use this as the numerator of a Z statistic. The appropriate denominator is the square root of the variance of A, under the null hypothesis of parity. This variance is sum (mrs/n(N-1))/(sum(m)^2), so the appropriate standardized statistic is

$$\text{weightedaveZ} = \text{sum}\ (\ m(a/m - c/n))/\text{sqrt}(\text{sum}(mrs/n(N-1))$$

This statistic can be written in terms of the within-cell Z's:

$$\text{weightedaveZ} = \text{sum}\ (wZ)/\text{sqrt}(\text{sumw}^2)$$

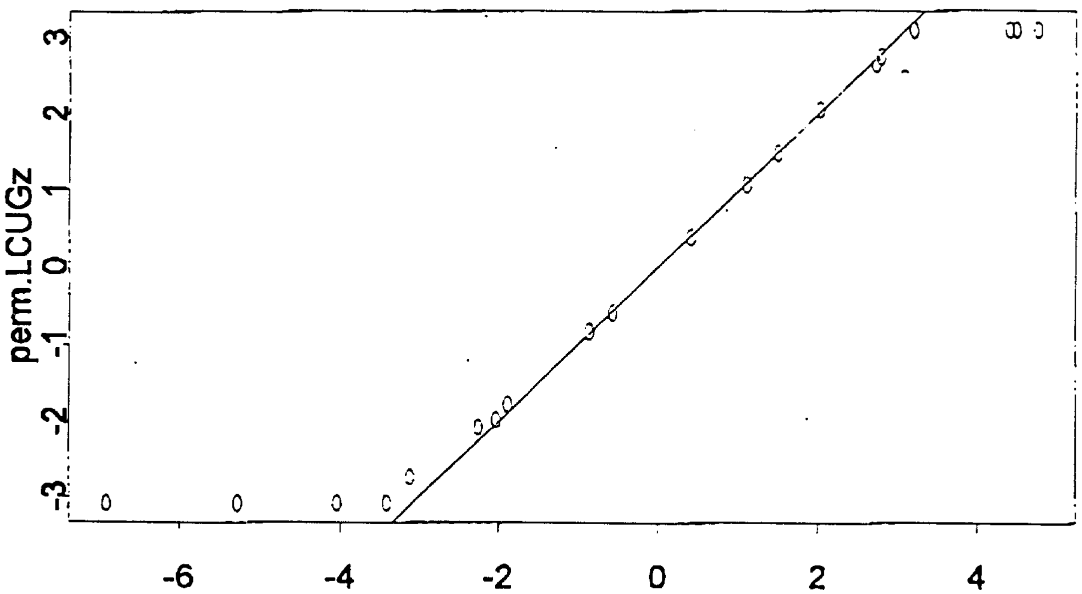
where the weight for each cell is

$$w = \text{sqrt}(mrs/n(N-1)).$$

In the Missed Installation analysis I used the more symmetric form

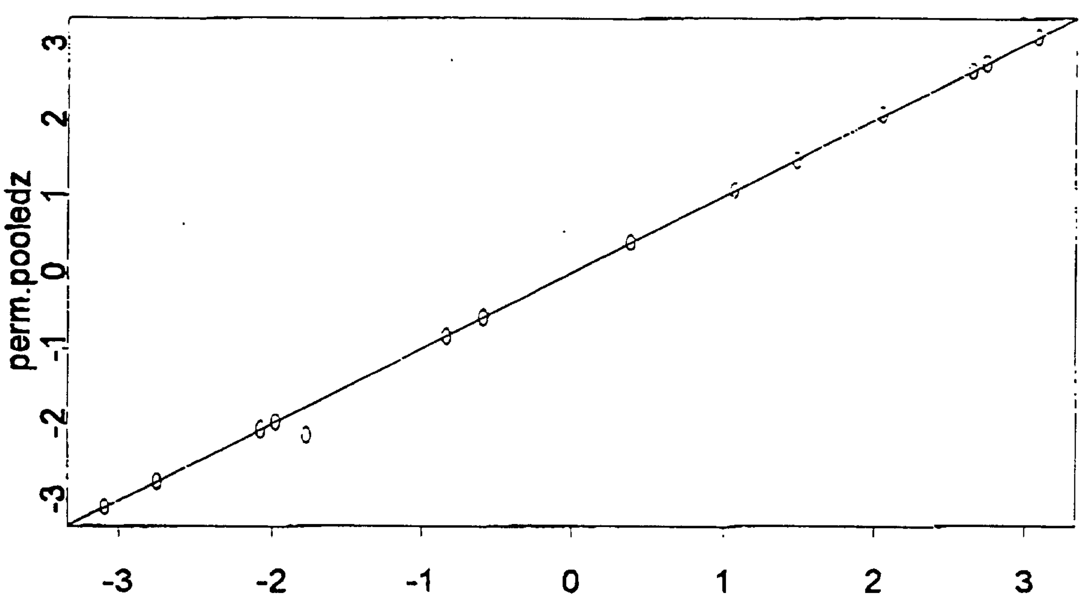
$$w = \text{sqrt}\ (rs/mn(N-1))$$

Figure 1. LCUGz and permLCUGz



LCUGz
cells with nCLEC >= 100

Figure 2. permLCUGz and permpooledz



perm.LCUGz
cells with nCLEC >= 100

Figure 3. LCUGz and permLCUGz

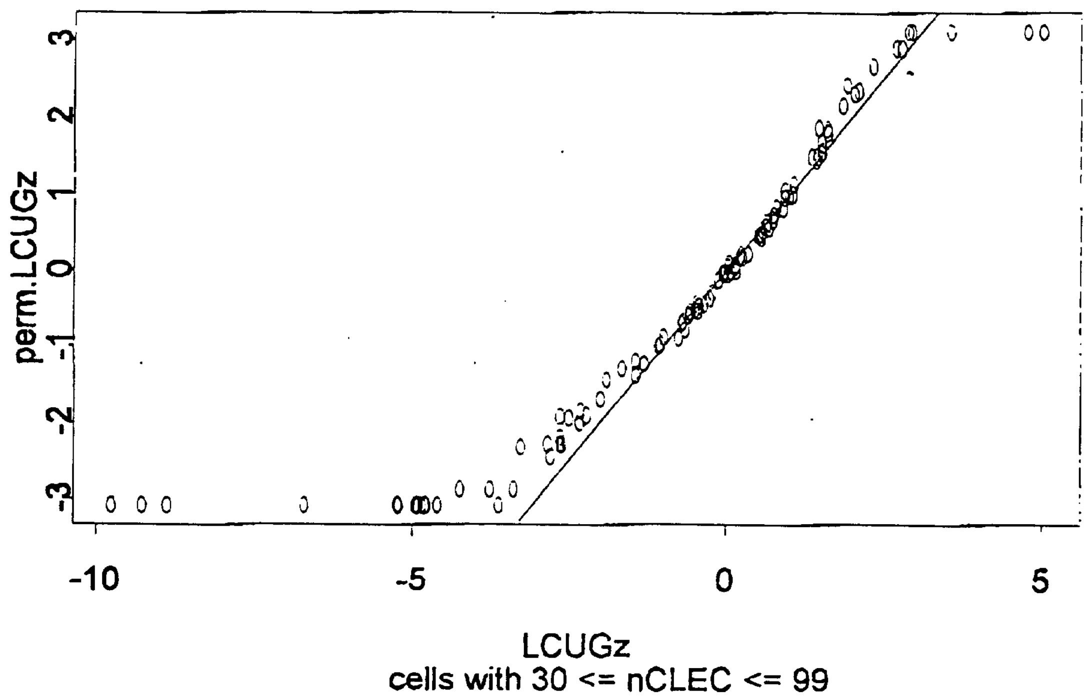


Figure 4. permLCUGz and permpooledz

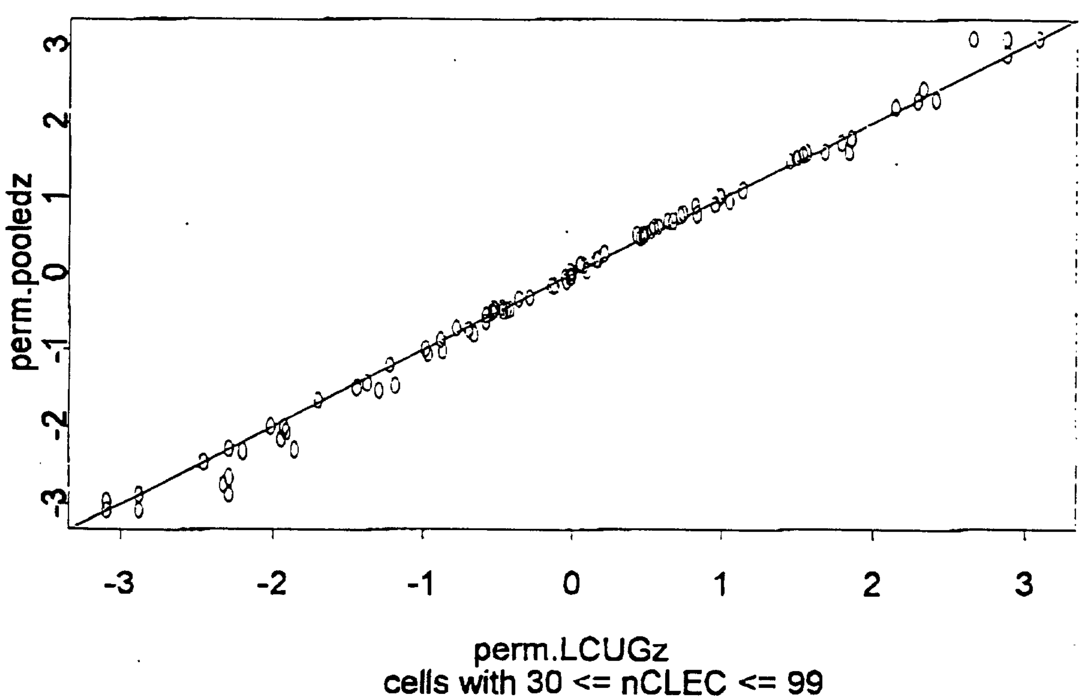
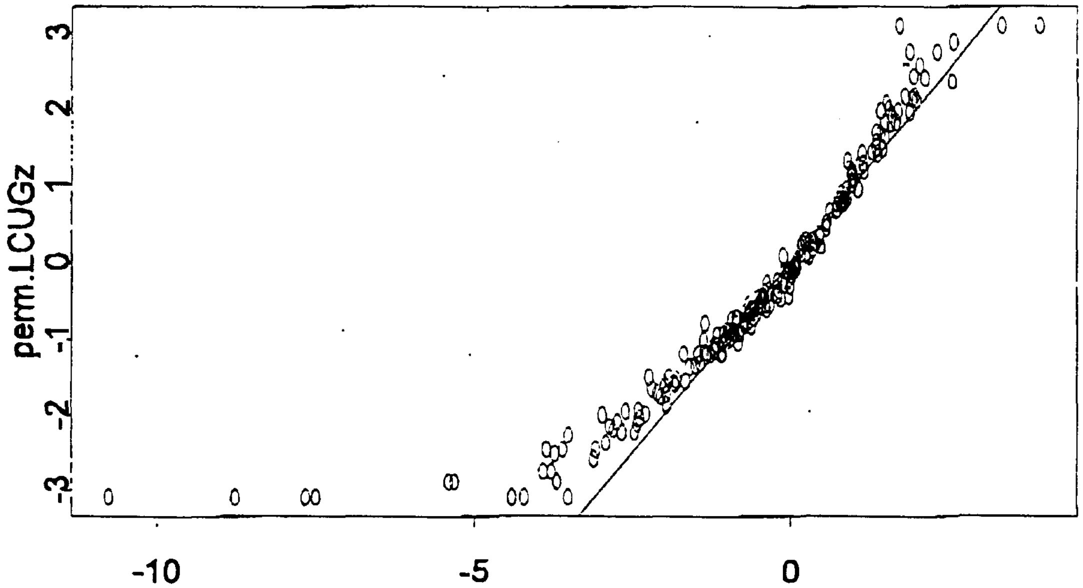
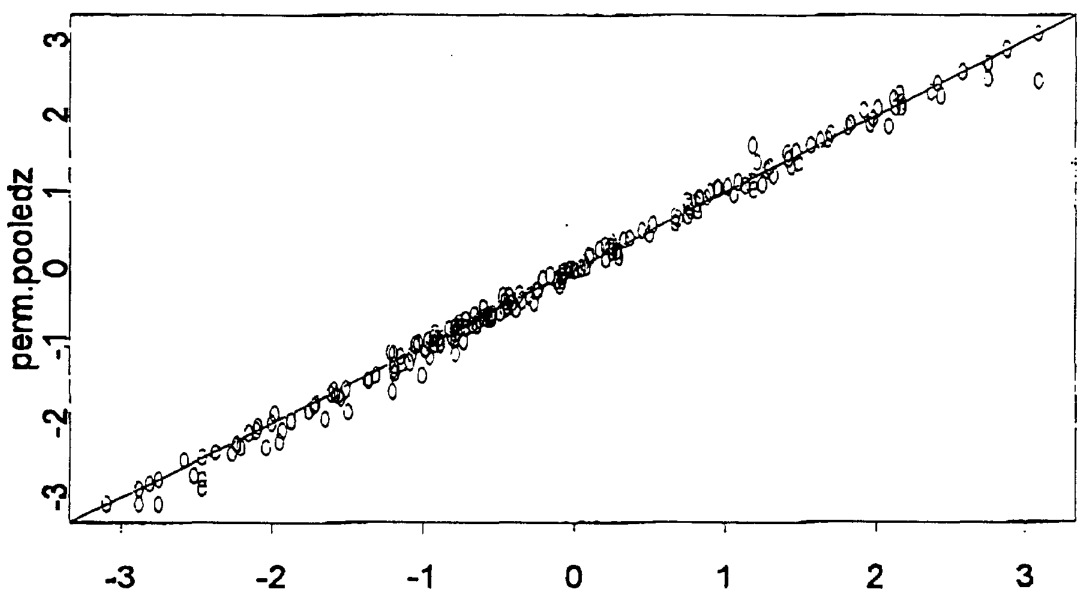


Figure 5. LCUGz and permLCUGz



LCUGz
cells with 10 <= nCLEC <= 29

Figure 6. permLCUGz and permpooledz



perm.LCUGz
cells with 10 <= nCLEC <= 29